UNCLASSIFIED

Defense Technical Information Center Compilation Part Notice

ADP010794

TITLE: Development of Virtual Auditory Interfaces

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: What is Essential for Virtual Reality Systems to Meet Human Performance Goals? [les Caracteristiques essentielles des systemes VR pour atteindre les objectifs militaires en matiere de performances humaines]

To order the complete compilation report, use: ADA390882

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, ect. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP010779 thru ADP010796

UNCLASSIFIED

Development of Virtual Auditory Interfaces

LCDR Russell D. Shilling[1], MSC, USN
United States Air Force Academy
Colorado Springs, CO 80840
USA

Tomasz Letowski, Ph.D.
Army Research Laboratory
Aberdeen Proving Ground, MD 21005
USA

1. Introduction

The design of visual components in virtual environments has shown rapid improvement and innovation. However, the design of auditory interfaces has lagged behind. Whereas visual scenes have become more compelling, the auditory portions of VE remain rudimentary. This disparity is perplexing since auditory cues play a crucial role in our day-to-day lives. Imagine entering a meeting with a room full of people. When you enter the room, you realize that the speaker's voice is emanating from all points in the room, yet the room is totally anechoic. In addition, you see other attendees moving in the room, yet there are no additional noises in the room except the speaker's voice. Despite walking into a "real" environment, your sense of reality would most probably be challenged. In fact, it is generally believed that the sense of presence is dependent upon auditory, visual, and tactile fidelity (Sheriden, 1996). Although the sense of realism in VE is also dependent on visual fidelity, virtual or spatial sound has been shown to increase the sense of "presence" (Hendrix, 1996). It stands to reason that when we develop poor auditory interfaces in a VE, the perceived quality of the entire VE is compromised (Storms, 1998). The problem with audio is that our normal auditory environment is "transparent". We don't consciously process a sound in our environment unless we NEED to attend to it. Yet, when slogging through mud while on patrol, soldiers use auditory cues to keep track of the people around them while scanning for threats in front of them. They don't need to keep looking at the people around them. While not consciously processing the sounds of their comrades, if someone stops walking, they'll recognize the lack of sound instantly.

2. Methods of Sound Presentation

There are a variety of ways to present sound in virtual environments. The most traditional method is to use speakers to present sound either monaurally, in stereo, or in surround sound. Speaker systems are bulky, do not typically provide elevation cues, and do not allow the sound engineer to have complete control of the auditory environment. Speaker systems DO allow for the possibility of presenting auditory stimuli such that the entire body is stimulated, especially when powerful subwoofers are employed. On the other hand, using headphones in conjunction with signal processing techniques, it is possible to generate stereo signals that contain most of the normal spatial cues available in the real world. Spatialized audio uses actual pinna cues stored as Head Related Transfer Functions (HRTFs) to give the perception of auditory objects as completely externalized in azimuth and elevation (Wightman & Kistler, 1989; Begault & Wenzel, 1993). When coupled with a headtracking device, spatialized audio provides a true virtual auditory interface. Using a spatialized auditory display, a variety of sound sources can be presented simultaneously at different directions and distances. One of the early criticisms of spatialized audio was that it was expensive to implement, however, as hardware and software solutions have proliferated, it has become feasible to include spatialized audio in most systems. Spatialized audio solutions can be fit into any budget, depending on the desired resolution and number of sound sources required. Most head-mounted displays are currently outfitted with headphones of sufficient quality to reproduce spatialized audio, making it relatively easy to incorporate spatialized audio in an immersive VR system. A complete lexicon for understanding and developing auditory displays can be found in Letowski, Vause, Shilling, Ballas, Brungert & McKinley (2000).

3. Effects of Auditory Displays on Performance

Illustrating the importance of sound, research conducted using spatialized auditory displays has demonstrated the importance of spatialized auditory cueing for reducing response time in cockpit applications. Spatialized auditory threat and attack displays were designed and implemented for both the pilot and co-pilot gunner in an AH-64 simulator at the Army Research Institute at Fort Rucker, Alabama (Shilling & Vause, 1999; Shilling, Letowski, & Storms, 2000). In this application, a ground-to-air missile display was supplemented with a spatialized auditory cue corresponding to the actual location of the missile relative to the pilot and co-pilot gunner. Figure 1 shows the difference between spatialized and normal displays for the response time to make the first 5 degrees of turn away from an incoming threat. Response time was reduced by approximately 350 msec. These data are consistent with previous research which demonstrated that response time to visual targets was significantly reduced when paired with a spatialized auditory stimulus (Perott et al., 1991) and the latency of saccadic eye movements was reduced when using spatialized auditory cues (Frens, Opstal & Willigen, 1995). In this same manner, auditory cueing can be used to compensate for the effects of limited FOV HMDs (Shilling, 1996). Applications can be further supplemented by exaggerating normal auditory cues

[1] For contact with author: Shilling85@alumni.wfu.edu

Paper presented at the RTO HFM Workshop on "What Is Essential for Virtual Reality Systems to Meet Military Human Performance Goals?", held in The Hague, The Netherlands, 13-15 April 2000, and published in RTO MP-058.

through so-called "supernormal localization" (Durlach, Shinn-Cunningham et al., 1993). Finally, using spatialized sound, speech intelligibility can be improved when applied to multi-user virtual environments and multi-channel radio communications (Haas, Gainer, Wightman, Couch & Shilling, 1997).

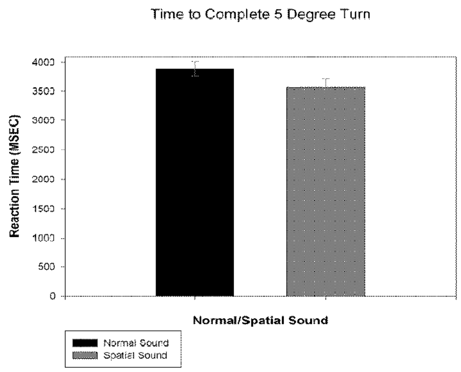

Figure 1: Difference between spatialized and normal displays

4. Lessons from the Entertainment Industry

The entertainment industry has recognized the importance of sound processing for over a century and has learned many important lessons that can be applied to problems in VE. At the beginning of the century, the Edison Standard Phonograph represented the cutting edge in audio technology. The method for cutting grooves in the wax cylinders was robust and resistant to the effects of scratches. However, consumers soon abandonedwax cylinders with vertically etched grooves for the less robust wax platter with horizontally etched grooves, because the platters were easier to store. Today, even though we have the technology to create astounding audio when developing VE's, it is more convenient to ignore the auditory interface because customer's aren't "requiring" high quality audio, software applications are not typically easy to implement, and the contributions of high quality sound are more subtle than for visual cues.

For instance, in motion pictures, sound has long been recognized as playing a crucial role in the emotional context of a film. Current efforts in my research are focusing on applying lessons learned from the film industry to problems associated with sound quality and emotional content in VE. Much can be learned about auditory special effects and sound system design from Hollywood. The first real attempt at immersing the audience in sound occurred with the production of Disney's "Fantasia" in 1939. Disney's sound engineers created a system called "Fantasound" which wrapped the musical compositions and sound effects of the movie around the audience. Though not a stereo production, the effects were quite astounding. However, the system required massive amounts of vacuum tube electronics and 54 speakers spread around the theater at a cost of $84,000 per theater. Virtually no theaters invested in the system and "Fantasound" was never used again. Today, we have a similar problem with applying sound in VE. Although the cost of consumer audio equipment has rapidly increased in quality and decreased in cost, systems designed for VE's are currently expensive and the development software to implement them is limited. Spatial audio sound servers, for example the AuSIM Acoustetron and the Tucker-Davis Technologies PD-1, typically cost in excess of $12,000. High cost and limited software availability are clearly the result of a lack of competition in audio products for VE.

5. Systematic Approach to Sound Design

On the practical side, the problem is not with the software engineers as much as with the lack of a clear set of requirements for implementing sound in VE. What is needed is a systematic approach to rendering the auditory environment necessary for any given application. When we want to render visual scenes, we rely on film as a reference. Unfortunately, when we design auditory scenes, we typically rely only on memory. In my laboratory, I am currently attempting to develop a systematic approach to cataloging the auditory environment to give the software engineer an objective reference to compare the sound in the VE with the real world experience.

One of the current efforts in my lab is to develop a systematic approach for obtaining baseline data concerning the content of an auditory environment. In addition to cataloguing the different sounds in a real environment, it is also important to systematically measure the intensity of sounds being experienced by the listener. In this manner, the VE developer has a highly detailed reference with which to compare the real world auditory environment with the virtual auditory environment. Two systems are currently being evaluated. The first system uses a portable Sony TCD-D8 DAT recorder coupled with Sennheisser microphone capsules (Figure 2). The microphone capsules will be inserted into an observer's auditory meatus (ear canal). In this manner, a complete spatialized recording can be made of the auditory environment, completely externalized with azimuth and elevation cues. The second system (Figure 3) is more robust, using a larger set of microphones produced by Core Sound which can clip to a set of eyeglasses to produce a binaural recording, complete with interaural time and intensity cues. Although, pinna cues cannot be utilized, the advantage of the latter system is that it would be more tolerant of extreme conditions, especially if the recordings are made outdoors. Both systems can be clipped to the belt and will be used in conjunction with a real time logging and event analyzer (CEL 593). The complete data set including sound recordings and sound measurements will be stored on CDROM for ease of use. The digital recordings also allow for spectral analyses to be conducted on specific auditory stimuli contained on the tape so that synthesized versions of those stimuli can be constructed.

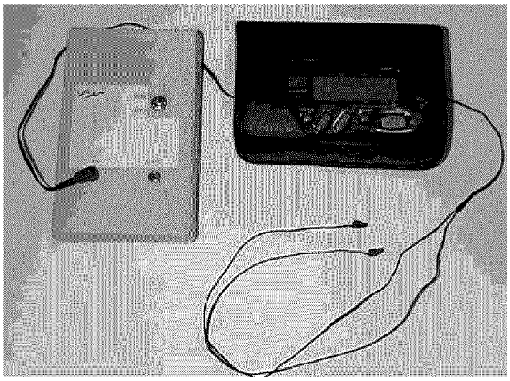

Figure 2: The used portable Sony TCD-D8 DAT recorder coupled with Sennheisser microphone capsules

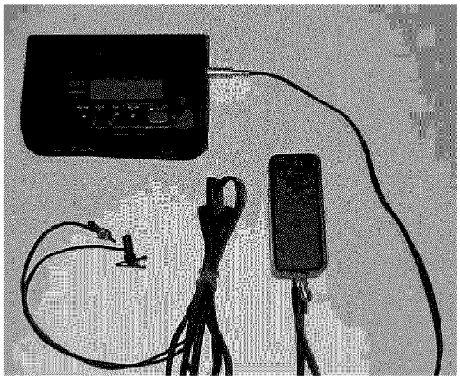

Figure 3: The set of microphones produced by Core Sound

References

Begault, D.R. & Wenzel, E.M. (1993). Headphone Localization of Speech. *Human Factors, 35 (2),* 361-376.

Durlach, N.I., Shinn-Cunningham, B.G. et al. (1993). Supernormal auditory localization. I. General background. *Presence, 2 (2),* 89-103.

Frens, M.A., Van Opstal, A.J. & Van der Willigen, R.F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Pereption & Psychophysics, 57 (6),* 802-816.

Haas, E. C., Gainer, C., Wightman, D., Couch, M. & Shilling, R.D. (1997). Enhancing System Safety with 3-D Audio Displays. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting.* 868-872, Albuquerque, NM.

Hendrix, C. & Barfield, W. (1996). The sense of presence in auditory virtual environments. *Presence, 5 (3),* 290-301.

Letowski, T., Vause, N., Shilling, R., Ballas, J., Brungert, D. & McKinley, R. (2000). *Human Factors Military Lexicon: Auditory Displays* (ARL Technical Report, ARL-TR-in print). Aberdeen Proving Ground, MD: Army Research Laboratory.

Perott, D.R., Sadralodabai, T., Saberi, K. & Strybel, T.Z. (1991). Aurally aided visual search in the central visual field: effects of visual load and visual enhancement of the target. *Human Factors, 33,* 389-400.

Sheriden, T.B. (1996). Further Musings on the Psychophysics of Presence. *Presence 5(2),* 241-246.

Shilling, R.D. & Vause, N. (1999). Tri-service cooperation, a key aspect of ARL-HRED spatial audio research. *Focus Army Research Laboratory, 5 (5).*

Shilling, R.D., Letowski, T. & Storms R. (2000). Spatial Auditory Displays for use within Attack Rotary Wing Aircraft. *Proceedings of the International Conference on Auditory Displays,* April 2000, Atlanta, GA.

Storms, R.L. (1998). *Auditory-visual cross-modal perception phenomena.* Doctoral Dissertation. Naval Postgraduate School, Monterey, California.

Wightman, F.L. & Kistler, D.J. (1989). Headphone simulation of free-field listening. II. Psychophysical validation. *Journal of the Acoustical Society of America, 85,* 868-878.

This page has been deliberately left blank

Page intentionnellement blanche